



An illustration of the use of an approach for treating model uncertainties in risk assessment

Torbjørn Bjerga, Terje Aven, Enrico Zio

► To cite this version:

Torbjørn Bjerga, Terje Aven, Enrico Zio. An illustration of the use of an approach for treating model uncertainties in risk assessment. Reliability Engineering and System Safety, 2014, 125, pp.46 - 53. 10.1016/j.ress.2014.01.014 . hal-01107605

HAL Id: hal-01107605

<https://hal.science/hal-01107605>

Submitted on 21 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An illustration of the use of an approach for treating model uncertainties in risk assessment

Torbjørn Bjerga^a, Terje Aven^a, Enrico Zio^{b,c}

^a University of Stavanger, Norway

^b Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy-Electricité de France, Ecole Centrale Paris and Supelec, France

^c Politecnico di Milano, Italy

Abstract:

This paper discusses an approach for treating model uncertainties in relation to quantitative risk assessments. The analysis is based on a conceptual framework where a distinction is made between *model error*—the difference between the model prediction and the true future quantity—and *model output uncertainty*—the (epistemic) uncertainty about the magnitude of this error. The aim of the paper is to provide further clarifications and explanations of important issues related to the understanding and implementation of the approach, using a detailed study of a Poisson model case as an illustration. Special focus is on the way the uncertainties are assessed.

1. INTRODUCTION

Quantitative risk assessment is largely based on models to represent systems and processes, and provide predictions of defined safety performance metrics. These models are conceptual constructs (translated into mathematical forms) built on a set of assumptions (hypotheses) on the systems and processes: for example, that the occurrence of an uncertain event of interest follows a Poisson distribution in time. The mathematical models include parameters, e.g., the (constant) rate of occurrence of the Poisson distribution: in practice, the values of these parameters are unknown and must be estimated, and the data and information available for that may be more or less precise and complete.

The modelling of a system or process needs to balance between two conflicting concerns: i) accurate representation of the phenomena and mechanisms in the system or process and ii) definition of the proper level of detail of the description of the phenomena and mechanisms so as to allow the timely and efficient use of the model. Differences between the real world quantities and the model outputs inevitably arise from the conflict of these two concerns.

In the field of risk assessment, probabilistic models are typically used to describe the (uncertain) future of the quantities characterising the system or process. From a theoretical viewpoint, these models reflect variation (often referred to as aleatory or stochastic uncertainty) in an infinite (large) population of elements similar to the one (those) studied (i.e. constitutively identical elements but behaving differently because of the intrinsic aleatory character of the properties which rule their behaviour). For example, the Poisson model referred to above can be used to represent the variation in the number of events occurring in a system over a period of time, when considering an (hypothetically infinite) population of similar systems. The values of the parameters of the probabilistic models need to be estimated from the information available on the system or process which, as said above, may be more or less precise and complete, and give rise to uncertainty on the values of the parameters (often referred to as epistemic uncertainty, meaning that it results from insufficient knowledge). As such, epistemic uncertainty can be reduced if additional knowledge and information can be acquired; on the contrary, aleatory uncertainty cannot be reduced and for this reason it is sometimes called irreducible uncertainty. The dichotomy of aleatory and epistemic uncertainty is instrumental for its treatment in risk assessment (Parry and Winter 1981, Apostolakis 1990, Helton 1994).

In synthesis, the models used in risk assessment are interpreted and simplified descriptions of the real systems and processes of interest, and their accuracy has to be balanced against their efficient use within the characteristic time scale of the assessment. As the value of a risk assessment is in providing informative support to decision making, the confidence that can be put in the accuracy, representativeness and completeness of the models is fundamental.

The concept of model uncertainty is pivotal in risk assessment and has been studied by several authors, see e.g. Zio and Apostolakis (1996), Devooght (1998), Nilsen and Aven (2003), Helton et al. (2004), Rosqvist and Tuominen (2004a, 2004b), Droguett and Mosleh (2008, 2013), Baraldi and Zio (2010), Vasseur et al. (2012), and Aven and Zio (2013), but there still lack consensus on how to treat it in practice and, even on the meaning to be given to it. It comes natural to address model uncertainty when there are alternative plausible hypotheses for modeling the specific phenomena or events of interest (Parry and Drouin 2009, Reinert and Apostolakis 2006), but it can also be evoked in relation to the difference between the actual values of the real world output and the values predicted by the model (Östergaard et al. 1996, Kaminski et al 2008, Nilsen and Aven 2003). Droguett and Mosleh (2008) also talk about model uncertainty in situations where

- a single model is generally accepted but not completely validated,
- a conceptually accepted and validated model is of uncertain quality of implementation,
- a single model covers only some and not all relevant aspects of the problem, and when
- composite models are formed by submodels of differing degrees of accuracy and credibility.

The problem of model uncertainty is, then, fundamental for the accreditation of the model itself, for its use in practice. We take the understanding of accreditation as the objective of reaching a required quality level of a model by validation, for its certified use. Clearly, this requires that model uncertainty be sufficiently small, for confidence in the use of the outputs produced by the model. What is sufficiently small is of course dependent on the purpose for which the model is to be used. In practice, model accreditation stands on the evaluation of the comparison of the model predictions with the corresponding true values of the predicted quantities, for establishing the level of confidence in the model predictive capability needed for the intended use of the model: the accreditation must demonstrate that in correspondence of given input values the model produces predictions of the true values of the output quantity with the sufficient level of accuracy and the confidence required for taking decisions. In the case of accreditation, then, the evaluation of the model uncertainty serves the purpose of verifying the level of accuracy achieved so as to have the confidence required to make decisions informed by the outcomes of the model.

In the case that experimental data are available, there exists a wide range of statistical methods that can be used for validation in order to accredit a model. These methods include both traditional statistical analysis and Bayesian procedures, see e.g. Bayarri et al. (2007), Jiang et al. (2009), Kennedy and Hagan (2001), Meeker and Escobar (1998), Xiong et al. (2009), and Zio (2006). However, these methods are not within the scope of the present paper, in which we consider situations with lack of data.

Model validation is often linked to model verification (and is often referred to as Verification and Validation, or simply “V&V.”), which is commonly understood as the process of comparing the model with specified requirements (Knupp 2002, McFarland 2008, Oberkampf and Trucano 2002, Rebba et al. 2006, Roache 1998). The verification part is obviously important in many contexts to produce a model that meets the specifications.

For the treatment of model uncertainty in the case of scarce data, both Bayesian and non-Bayesian approaches can be undertaken. Classic examples of Bayesian approaches are the alternate hypotheses and adjustment factor approaches (Zio and Apostolakis, 1996). In the former, alternate hypotheses approach, a plausible set of models based on alternate hypotheses are used. These hypotheses are then assigned individual probabilities reflecting the analyst’s relative confidence in the alternate hypotheses to be true. Differently, the adjustment-factor approach uses the output of a single-best model which is then adjusted by a multiplicative or additive factor to account for the uncertainty directly. Since this factor is in general unknown, probability distributions are introduced to provide a measure of confidence for different values of these factors. As for non-Bayesian approaches, an example is that of Rosqvist and Tuominen (2004a) which is based on a qualitative score assessment of direction of bias toward risk, where each modeling assumption is given a score: no bias, conservative or optimistic. For instance, if an assumption is deemed to represent the physical or social phenomena truthfully without any bias, then it is given the score ‘no bias’.

One structured way for addressing the problem of model uncertainty in risk analysis is that described in the NUREG 1855 report (issued in 2009 by the US Nuclear Regulatory Commission (a draft of a revised version open for

comments was issued in March 2013; the draft is basically identical to the 2009 version on the issues here discussed) and related documents. What we find in the report mainly relates to the sources of model uncertainty (NUREG 1855, p. 14):

Model uncertainty arises because different approaches may exist to represent certain aspects of plant response and none is clearly more correct than another. Uncertainty with regard to the PRA results are then introduced because uncertainty exists with regard to which model appropriately represents that aspect of the plant being modeled. In addition, a model may not be available to represent a particular aspect of the plant. Uncertainty with regard to the PRA results is again introduced because there is uncertainty with regard to a potentially significant contributor not being considered in the PRA.

The statements quoted above are not sufficiently unambiguous to confidently treat model uncertainty in risk assessment. The sentence “uncertainty exists with regard to which model appropriately represents that aspect of the plant being modeled”, in practice is not necessarily meaningful for all cases. If you are considering two models, one model may give more accurate output for some inputs, and the other for others: what is, then, the model that “appropriately represents those aspects ... being modeled”? And what is the uncertainty about it?

In another part of the NUREG 1855 report (p. 7), it reads:

In developing the sources of model uncertainty, a model uncertainty needs to be distinguished from an assumption or approximation that is made, for example, on the needed level of detail. Although these assumptions and approximations can influence the decision making, they are generally not considered to be model uncertainties because the level of detail in the PRA model could be enhanced, if necessary. Therefore, methods for addressing this aspect are not explicitly included in this report, and Section 5 discusses their consideration.

It seems, then, that assumptions and approximations on the level of resolution of the model are not included as contributors to model uncertainty. On the contrary, we believe that they are key contributors and need to be included in the analysis.

The present paper is based on the framework for model uncertainty analysis introduced by Aven and Zio (2013) with the aim of clarifying how to interpret and treat model uncertainty. From the above discussion it appears that such clarifications could be important for the risk field, and risk regulation in particular. In Bjerga et al. (2012), an application of the framework has been presented within a risk assessment related to hydrocarbon releases in an LNG (Liquefied Natural Gas) plant in an urban area. Here, we extend the work by considering a probability model (the Poisson model) for describing the variation in the occurrences in time of a specific event. Through the example, we clarify the meaning of the various concepts of the model uncertainty framework and show how they can be described and measured in practice using different approaches, including subjective probabilities and interval probabilities. We also deal with the decision on accreditations or remodeling. Before we introduce the Poisson model and study its uncertainties, we give a short presentation recall of the framework. To help the reader understand the concepts used this case study and the framework in general, we first provide some fundamentals related to probability models.

2. SOME FUNDAMENTALS RELATED TO PROBABILITY MODELS

Let N be the number of events occurring stochastically in a specific system for an interval of length 1, and assume that the probability of the values that N takes can be described by a Poisson distribution with expected value λ . We refer to this distribution as a (probability) model with parameter λ . As a model, it represents a (probabilistic) description of a phenomenon of the real world – it is an approximated description of how this phenomenon occurs. In statistical analyses (whether traditional or Bayesian), its parameter λ has a “true” value. Now, what is the interpretation of this “true” value? Does it, in fact, make sense at all to talk of “true” value of the model parameter?

A probability model is based on a set-up which is thought-constructed and refers to an infinite number of instances similar to the one under study. In the case investigated here (Section 4), the model refers to an infinite population of similar systems for describing the variation in the number of events occurring in these systems for the specified

interval of time. What “similar” means has to be clarified - in a Bayesian context it means that the numbers in different systems can be judged exchangeable (a sequence of random quantities is exchangeable if their joint probability distributions are independent of the order of the quantities in the sequence), whereas in a traditional statistical context similar means that the N s are independent and identically distributed.

In this modelling set-up, one can refer to a “true” distribution F describing the variation in the number of events occurring in the infinite population of similar systems. We write true in quotes because its meaning exists only within this thought-constructed set-up. As a model of this “true” distribution, we introduce the Poisson model. The arguments for why this model is suitable to describe the situation of interest are well-known from text-books in probability theory. By extension of this reasoning, we can talk of a “true” value of the parameter λ , to be interpreted as the average number of events in the infinite population of similar systems. Both F and λ are unknown and must be estimated. In a Bayesian analysis, focus is on the epistemic uncertainties about this “true” value of λ (expressed as prior and posterior distributions). In a traditional statistical analysis, one tries to estimate the “true” value of λ and give a confidence interval for it (for the sake of simplicity, it is common to say that one estimates λ and makes confidence intervals for λ). Normally, less attention is paid to the epistemic uncertainties related to the probability model itself, but from the above reasoning it appears that it make sense to address it in relation to the deviation between the “true” F and Poisson distributions. For readability we will in the following avoid writing true in quotes.

3. THE FRAMEWORK FOR ANALYSING MODEL UNCERTAINTY

We consider an event/system/process subject to a risk assessment, and assume that at the time of the assessment no experimental data are available. Consider a quantity Z whose true value is realized in the *future*, and which we are interested in knowing. As an example, Z could be the actual number of fatalities due to a potential outbreak of a new virus. The actual value of Z cannot be known till after a potential outbreak. To predict the future value of Z a model $G(X)$ is developed. Both X and Z may be vectors. A simple model would be $G(X) = G(X_1, X_2) = X_1 X_2$, where X_1 is the fatality rate and X_2 is the number of exposed people (say the number of citizens in a country). The predictions by $G(X)$, then, depend on the structure G and the parameters X_1 and X_2 .

Define:

Model error: The difference, $\Delta G(X)$, between the model prediction $G(X)$ and the true future value Z , i.e. $\Delta G(X) = G(X) - Z$.

Model output uncertainty: Uncertainty about the magnitude of the model error.

Note that according to this definition, model error and model output uncertainty are different, yet connected, concepts. The distance between the present time prediction $G(X)$ and the future value Z is a well-defined quantity, referred to as the model error. As this model error cannot be known at the time of the prediction, we have uncertainty – namely, model output uncertainty. Returning to the above virus example, say that prior to the outbreak of a virus the model $G(X)$ predicts 50 fatalities; then, after the outbreak is over we count 33 fatalities caused by the virus. The derived distance, $50-33=17$, is the true model error. However, at the time of the prediction we cannot know with certainty that the true model error is 17: we are facing model output uncertainty. This uncertainty is actually *epistemic* uncertainty about the value of the model error and hence it may in theory be assessed using a suitable tool for measuring this type of uncertainty, like subjective probabilities and interval probabilities.

In line with Aven and Zio (2013), model output uncertainty results from two components:

Structural model uncertainty: The conditional uncertainty associated to the model error $\Delta_G(X)$, given the true value X_{True} (i.e. $\Delta_G(X_{\text{True}})$).

Input quantity (parameter) uncertainty: The uncertainty associated with the true value of the input quantity X .

The structural model uncertainty is expressing the epistemic uncertainty under the condition that the input parameters are known (the true values). In other words, the structural model uncertainty expresses uncertainty about the model error when we can ignore uncertainty about the parameters X , and relates then to the model structure G itself. Typically this uncertainty is associated with assumptions and suppositions, approximations and simplifications made in the modeling. Input quantity (parameter) uncertainty is on the other hand reflecting epistemic uncertainties about the true value of X . The meaning of true value is described for probability models in Section 2.

Sources of structural model uncertainty stem from actual “gaps” in knowledge which can take the form of poor understanding of phenomena that are known to occur in the system, as well as complete ignorance of other phenomena. This type of uncertainty can lead to “erroneous” assumptions regarding the model structure. Other sources of structural model uncertainty stem from approximations and simplifications introduced in order to translate the conceptual models into tractable mathematical expressions.

The framework proposed in Aven and Zio (2013) links the objectives of modeling and risk assessment (cf. de Rocquigny et al. 2008) to model output uncertainty, and points at uncertainty analysis as a tool to accredit a model so as to ensure a certain quality and possible certification. In the accreditation process, the understanding of the influence of uncertainties on the results of the analysis is of importance, to adequately guide the uncertainty reductions. If the model considered cannot be accredited, remodeling is required. When an accredited model is obtained, a risk analysis might be conducted to inform the decision makers on the selection and compliance in line with the objectives stated above.

The characteristic that no experimental data exist at the time of the assessment leads us away from classical statistical tools for validation and subsequent accreditation of the model. Instead, validation transforms into utilizing expert/analyst argumentation based on established scientific theories and specific knowledge about the system, which the model assessed, intends to describe.

An important observation is that no restrictions pertain to utilizing a pure probability-based approach. The framework opens for both probabilistic and non-probabilistic approaches, and thereby injects flexibility into the uncertainty analysis, giving the opportunity to choose the approach that is judged to best represent/express the uncertainties, given the specific phenomena and surroundings examined.

4. CASE STUDY: A POISSON MODEL

The case study pertains to a risk assessment context for a specific activity, where the modeling of the uncertain occurrences in the future of a type of undesirable events are described by a Poisson model. Let $N(t)$ be the number of events occurring in the time interval $[0, t]$. It is assumed that the stochastic process N is a homogeneous Poisson process with occurrence rate λ . Hence $N(t)$ has Poisson distribution with expected value λt , i.e. $P(N(t) = n | \lambda, t) = p(n | \lambda, t) = (\lambda t)^n e^{-\lambda t} / n!$, $n = 0, 1, 2, \dots$. We interpret λ as the expected number of events occurring per unit of time.

Furthermore let $p_0(n | t)$ be the true distribution of the number of events in $[0, t]$, obtained by considering an infinite number of activities similar to the one considered. The average number of events per unit time is defined as λ_0 . The Poisson distribution $p(n | \lambda, t)$ is a model of this true distribution, and λ_0 is the (unknown) true value of the model parameter λ . The distributions p_0 and p represent the true variation in the number of events occurring in such intervals and the variation as modeled, respectively.

In the case study, the objective of the risk assessment is to verify that the 0.95 quantile, n_{95} , of $p_0(n | t_0)$ is in compliance with a regulatory threshold value n_M , where t_0 is a fixed point in time.

In line with the framework presented in Section 3, we identify n_{95} as the quantity of interest, Z , λ as the parameter X , and the model representing Z , $G(\lambda)$ as the 0.95 quantile of the Poisson distribution, which we refer to as $n_{95}(\lambda)$.

The model error can thus be written $\Delta G(\lambda) = G(\lambda) - n_{95} = n_{95}(\lambda) - n_{95}$. The structural model uncertainty relates to uncertainty about the value of $\Delta G(\lambda_0) = G(\lambda_0) - n_{95} = n_{95}(\lambda_0) - n_{95}$, and the parameter uncertainty to the true value of λ , i.e. λ_0 .

As a concrete example of this setting, we can consider potential releases from a planned commercial pilot facility/system handling crude oil with new technology. The system operates in a seasonal market following economic cycles and variations in demand. The project period is scheduled for five years. The governing authorities (agency) urge that compliance with relevant regulations be demonstrated prior to construction. Concerning the environmental risk and potential releases, the authorities acknowledge that releases could occur due to the novel technology and the limited operational experience. The authorities have specified an acceptance level of 5 releases during one month, to license construction and continuous operation; the system must be demonstrated capable of holding this true with a probability of 0.95.

For the model of the number of releases, a homogeneous Poisson process is initially found to be representative. The parameter λ representing the average number of releases is estimated to be 1.75 per month, based on an analysis of the technical solutions at the facility. The 0.95 quantile is calculated to $n_{95}(1.75) \approx 4$, and it is concluded that the requirement from the authorities is met (the probability of having more than 5 releases is calculated to be approximately 0.01).

But what about uncertainties, and model uncertainties in particular? In the following we will address this issue by conducting an uncertainty analysis with different approaches and perspectives, all in line with the structure and terminology presented in Section 3.

4.1. Uncertainty Analysis

The uncertainty analysis is conducted by a group of experts, and to evaluate model output uncertainty and conclude on model accreditation various approaches are applicable. The group chooses the following three approaches:

- i) A *qualitative scheme* giving scores on the importance of the assumptions made, reflecting both the degree of sensitivity and the uncertainty (see Flage and Aven 2009, Selvik and Aven 2011). A guideline for assigning scores is outlined in Appendix A.
- ii) *Subjective probabilities*: if we assign a probability of 0.1, say, it means that the assigner has the same uncertainty or degree of belief for this event to occur as for drawing a specific ball out of an urn containing 10 balls (Lindley 2000, 2006). If T is an unknown quantity taking values on the real line, a subjective probability distribution $F(t)$ can be expressed so that $F(t) = P(T \leq t)$, for all t , and being interpreted with reference to the urn standard. A subjective probability is always conditional on some background knowledge, but this is often suppressed in writing to avoid unnecessarily complicated notation.
- iii) *Imprecise probabilities* where assigning an imprecision interval, e.g. $[0.1, 0.2]$ for a probability $P(A)$, means that the assigner states that his/her degree of belief for an event A to occur is greater than or equal to the urn chance of 0.10 (the degree of belief of drawing a specific ball out of an urn containing 10 balls) and less than or equal to the urn chance of 0.20. The analyst is not willing to make any further judgments. A subjective probability of an event A is a special case of an imprecise probability, where the interval boundaries coincide.

4.1.1 Using a qualitative scheme

First the Poisson model is scrutinized by studying the underlying model conditions against sources of structural model uncertainty. The homogeneous Poisson process can be defined as follows:

A stochastic process N is a homogeneous Poisson process with rate λ if,
 $E[N(t+h) - N(t) \mid \text{history up to } t]/h$ converges to λ , as h converges to zero, for all t .

From the definition, we may identify two critical model conditions for the Poisson process:

1. *Independence*: The occurrence rate for the events at a specific point in time t is independent of the history up to that time.
2. *Stationarity*: The occurrence rate is a constant, λ .

The analyst group discusses the uncertainties related to these conditions. Several sources of uncertainties in this regard are pointed at, here formulated as assumptions:

- a) There is no deterioration of the system over time.
- b) A release will not cause learning effects.
- c) Utilization of the system does not vary in time – it is assumed to be 50% of full capacity.
- d) If a release occurs, the chance of new releases close in time is not increasing.

In line with the procedure of approach i), a qualitative importance score is assigned for each of the assumptions justifying the model. The first step in the approach is a systematic identification of the assumptions, and is done by the risk analyst. An independent review should be performed to ensure that all key assumptions are identified. The sources of uncertainty a) through d) cover the list of assumptions in this case. Next, a classification of the uncertainties and sensitivities in three categories (high, medium and low) is performed following the guideline in Appendix A. To obtain a high importance score, the assumptions must be subject to large uncertainties, and the model conclusions must be sensitive to changes in the assumptions. The score evaluations are presented in Table 1, followed by brief explanations in the subsequent paragraphs:

Table 1. Importance score for assumptions a) through d).

Assumption	Uncertainty	Sensitivity	Importance
a)	Low	Medium	Low-Medium
b)	Medium	Medium	Medium
c)	High	High	High
d)	Low	Low	Low

a) Over time the system will experience deterioration (causing e.g. material fatigue and rupture), which is increasing the chance of releases over time. On the other hand, maintenance plans are prepared to restrain the system degradation. Accordingly the analyst group rates the uncertainty as low. The assumption is considered medium sensitive, as relatively much deterioration is required to violate the model conditions.

b) In case of a release it is likely that there will be a focus on causation, to be able to understand what happened and give input to how to prevent similar releases in the future. The occurrence rate thus depends to some extent on its history. The uncertainty about this assumption is judged medium. The sensitivity is also set to medium; considerable learning effect is necessary to alter the model choice.

c) There are also concerns related to the utilization of the system, which is assumed responsive to shifting demand and economic cycles. For instance it is questioned whether the utilization via demand is correlated with seasonal/volatile oil prices; i.e. high prices compel high utilization, hence a higher chance of releases, and vice versa. The group finds the assumption of fixed utilization as a strong simplification subject to large uncertainties, and the stationary assumption (2), in particular, is found sensitive to changes.

d) Lastly, given a release, then a second release is more likely to occur close in time due to extra stress imposed on the system. However the group finds that the additional stress is minor and assigns low scores to this assumption.

Given the purpose of the risk assessment, to demonstrate that the acceptable rate of 5 releases per month will not be overcome with probability 0.95, the analyst group considers assumptions a) and d) to be acceptable; the effect on the model error is judged small enough for accreditation. The group finds the medium importance score for assumption b) to be on the line of acceptable/unacceptable. By arguing that any learning effect will cause a reduction of releases-no learning is the worst case scenario-the group concludes it to be acceptable. The importance

score ‘high’ on assumption c) suggests that the homogeneous Poisson model is invalidated and remodeling should be seriously considered. The remodeling should relate to the form of the occurrence rate.

Secondly, a parameter uncertainty assessment is carried out. The parameter uncertainties are addressed in different ways. Using standard Bayesian statistical procedures, for example, the 95% credibility interval [0.5, 3.0] for λ is established. The meaning of this interval is that with (subjective) probability 0.95 it covers the true value of λ .

4.1.2 Using subjective probabilities

Next a quantitative analysis according to approach ii) is conducted, using subjective probabilities. We take lessons from the results of the above crude qualitative approach (i) and here focus only on assumption c). To quantify the structural model error $\Delta G(\lambda_0)$ due to assumption c) means to compare n_{95} with $n_{95}(\lambda)$ for various (true) values of λ , where n_{95} represents the true (unknown) 0.95 quantile and $n_{95}(\lambda)$ is the 0.95 quantile when using the Poisson model with the (true) parameter λ .

Figure 1 below shows $n_{95}(\lambda)$ for $\lambda = 0.5, 1.75$ and 3 . We see that $n_{95}(0.5) = 2$, $n_{95}(1.75) = 4$ and $n_{95}(3) = 6$. The Figure is simply depicting the fact that the values 2, 4 and 6 are the 0.95 quantiles of the Poisson distribution when the parameter is 0.5, 1.75 and 3, respectively.

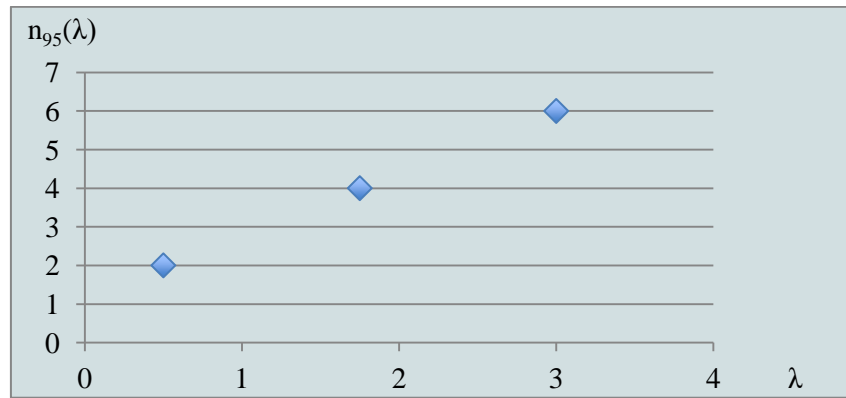


Figure 1. The Poisson 0.95 quantile $n_{95}(\lambda)$ for $\lambda = 0.5, 1.75$ and 3 .

To illustrate the analysis, let us consider the case of $\lambda = 1.75$. This means that we look into the structural model uncertainty about the structural model error $\Delta G(1.75) = n_{95}(1.75) - n_{95} = 4 - n_{95}$. Since the true 0.95 quantile, n_{95} , is unknown at the present time, we need to perform an assessment. First, we establish a credible outcome space. The group does this by linking n_{95} to the number of releases per month, which can be equal to, or greater than zero. The analysts group makes the judgment that more than 10 releases in a month are so unlikely that this eventuality can be ignored. Hence, the true 0.95 quantile value, n_{95} , is judged to lie in the interval $[0, 10]$. Then, the interval $[-6, 4]$ releases per month is the range of the structural model error outcome for $\Delta G(1.75)$.

Having established the structural model error outcome range $[-6, 4]$ for $\Delta G(1.75)$, the experts group proceeds with assigning subjective probabilities to the values therein. Clustering the outcomes is found useful and a triangular distribution results, as shown in Figure 2. For example, from Figure 2 we see that the structural model error is in the cluster $\{-1, 0, 1\}$ (i.e., the true 0.95 quantile value is either 3, 4 or 5) with probability 0.4. In the following paragraph, we explain further the group reasoning leading to such distribution.

We condition on perfect knowledge about the expected release rate, here equal to 1.75. However, this is the average in the long run, and we could experience a process where the actual 0.95 quantile n_{95} deviates from 4 due to assumption c). We may have a situation where the occurrence rate is low in some periods and high in others, still giving an average of 1.75, but resulting in 0.95 quantile values different than 4. Then, assumption c) leads to the rather broad distribution of Figure 2, and the homogeneous Poisson model cannot be accepted.

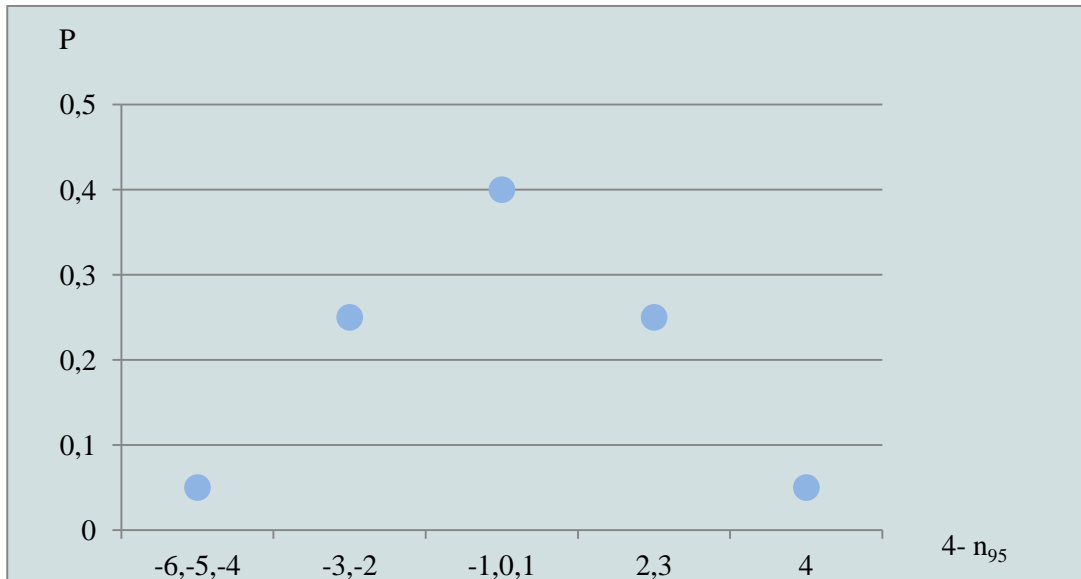


Figure 2. Subjective probabilities for $4 - n_{95}$ due to assumption c)

4.1.3. Using imprecision intervals

An attempt was also carried out to assign imprecision intervals for the structural model output uncertainty in line with approach iii). The group again focused on assumption c), adopted the same occurrence rate, $\lambda = 1.75$, and used the same structural model error outcome range $[-6, 4]$ as with approach ii). The results are displayed in Figure 3. As a concrete example, consider the structural model error of $\{-1, 0, 1\}$ (the true 0.95 quantile is either 3, 4 or 5); the interval displayed expresses that the analysts' subjective probability for the model error to be $\{-1, 0, 1\}$, is between 0.2 and 0.5. The analysts are not willing to make more precise assignments given their knowledge. The justification of the numbers follow the same lines as for approach ii), but the group quickly concluded that it was difficult to determine any concrete interval – specifying one number was difficult, but two was nearly unreachable.

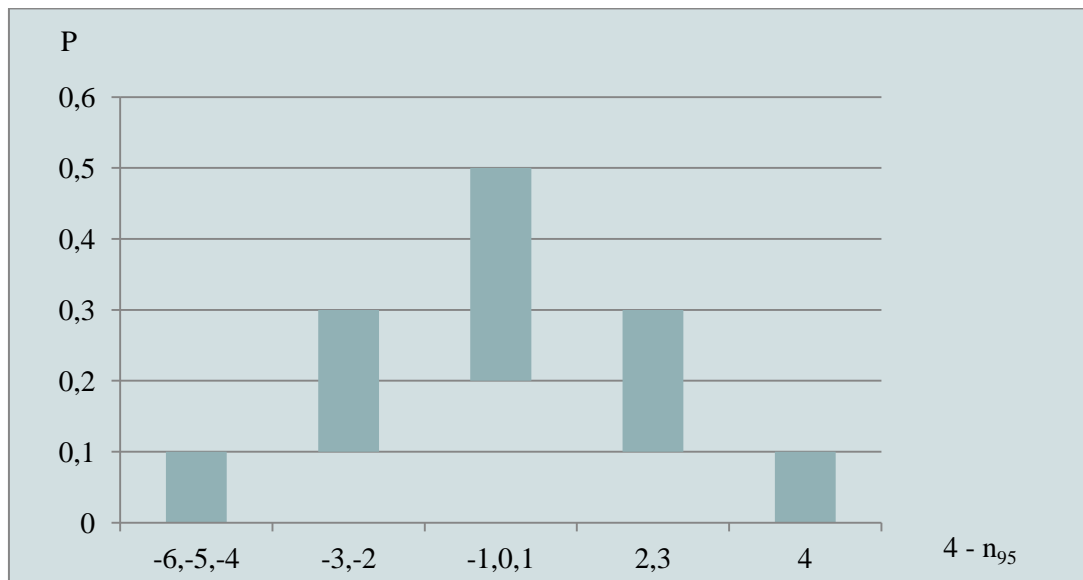


Figure 3. Imprecision intervals for $4 - n_{95}$ due to assumption c).

4.1.4 Results

Based on this analysis and seen in relation with the model purpose, the analyst group concludes that the uncertainties associated with the homogeneous Poisson model is not acceptable: the assumption c) requires remodeling.

4.2. Remodeling

The analysis showed that the stationary assumption is violated as a consequence of the uncertainty related to assumption c). This information suggests that a Non-Homogeneous Poisson Process (NHPP), which relaxes the stationary requirement, may be a suitable candidate. Starting from this model we repeat the above verification process and assess the model output uncertainties with approaches i), ii) and iii).

A non-homogeneous Poisson process differs from a homogeneous Poisson process in replacing the constant, λ , with an intensity function varying with time, $\lambda(t)$. The expected number of events in a given time interval, $[a,b]$ is then $\int_a^b \lambda(t)dt$. The critical model conditions are:

1. Independence: That the occurrence rate for the events at a specific point in time t is independent of the history up to that time.
2. The occurrence rate is a deterministic function $\lambda(t)$ varying with time.

The first assumption is the same as for the homogeneous process. The second assumption brings up the question on how to specify the intensity function $\lambda(t)$. The analysis group believes that the utilization is linked to the oil price, and a quick search into historical prices reveals a well-entrenched trend, high prices during summer versus low prices during winter caused by the underlying seasonal car-driving pattern. Based on the price trend found, held up against anticipated utilization, the group proposes an intensity function that varies from 0.5 up to 3 releases per month specified as follows:

$$\lambda(t) = \begin{cases} 0.5 & \text{if } t \in [\text{Dec} - \text{Feb}] \\ 1.75 & \text{if } t \in [\text{Mar} - \text{Jul}, \text{Nov}] \\ 3 & \text{if } t \in [\text{Aug} - \text{Oct}] \end{cases}$$

This implies that for each of the months Aug – Oct there is 0.08 probability for exceeding the limit of 5 releases per month, and less than 0.01 probability the remainder of the year. This in turn means that risk reducing measures for the pilot plant should be implemented to comply with the regulatory 0.95 probability criterion.

In the list covering the assumptions a) through d), c) is replaced with an assumption covering the enhanced knowledge level:

c’): Utilization of the system *varies with time/oil prices* causing the release rate to vary according to $\lambda(t)$.

The assumption still has associated uncertainty, which needs to be addressed. Firstly, it is recognized that the seasonal pattern is based on historical data. The amount of data is substantial and showing little deviation between years, and the group possesses no knowledge suggesting this would change in the years to come. Secondly, the resolution is very crude compared to the observed oil prices, but in practice some inertia in the system will exist, e.g. due to the fact that some storage tanks are part of the production system. Thirdly, there could be other factors influencing the utilization besides oil prices, as this is a novel product. These factors lead the group to assign a medium score to assumption c’), as seen from Table 3.

Table 3. Importance Score for Assumptions a), b), c’) and d).

Assumption	Uncertainty	Sensitivity	Importance
a)	Low	Medium	Low-Medium
b)	Medium	Medium	Medium
c’)	Medium	Medium	Medium
d)	Low	Low	Low

The medium score seen in relation with the purpose of the risk assessment resulted in the model being accredited. A similar analysis based on subjective probabilities and intervals was carried out as in the previous Section. The distributions obtained have a narrower shape than those in Figures 1 and 2, and it is concluded that the NHPP is acceptable. The details are omitted as the analysis is conceptually analogous to the one carried out above.

Consequently the NHPP with the specified deterministic intensity function $\lambda(t)$ is accredited for use in assessing the number of releases from the pilot facility. As noted earlier this will imply that authority limits are breached in the months Aug-Oct, and risk reducing measures could be necessary; possibly a restrained production rate in the mentioned months. To find a suitable measure is however not the objective in this paper and will not be elaborated on any further.

5. DISCUSSION

The motivation and use of the framework studied in the present paper is thoroughly discussed in Aven and Zio (2013). Of the many issues there raised, we emphasize one key point: model output uncertainty is not the same as the model error $\Delta_G(X)$; it is actually the epistemic uncertainty about its magnitude. To measure this uncertainty different tool can be used as illustrated above. The model output uncertainty and its measurement are considered in relation to the magnitude of the model error, as is clear from the analysis in Section 4. Hence, what is a small model output uncertainty cannot be seen in isolation from the model error.

From Section 4, analysis and treatment of model uncertainties are based on the algorithm seen in Figure 4:

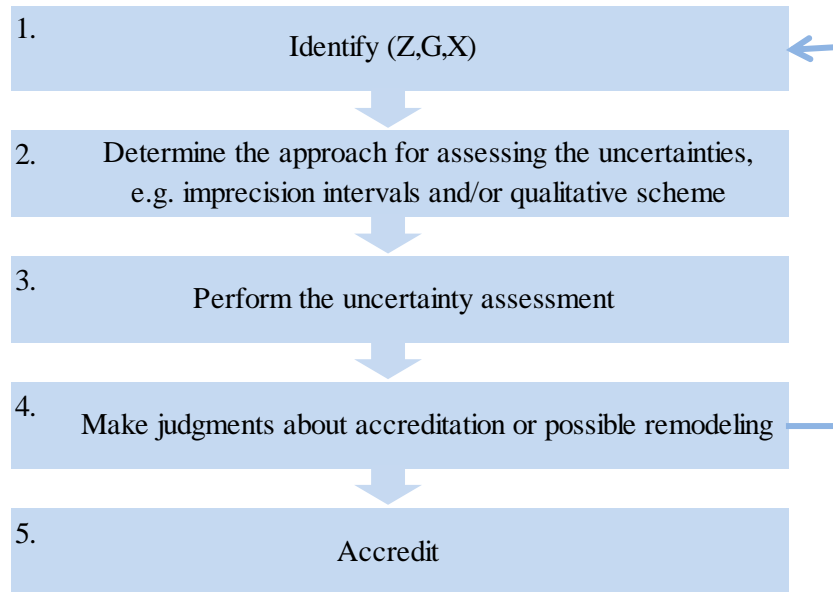


Fig. 4. Steps in the analysis and treatment of model uncertainty.

A key issue to determine in relation to step 1 is whether the quantities of interest Z and X are probabilistic parameters of probability models (as in the example considered in Section 4) or some physical quantities, e.g. the number of events occurring or the number of fatalities as studied in Bjerga et al. (2012). For proper analysis of the model uncertainties, precision is required as G is strongly dependent on this.

For the choice of approach in step 2, it is clear that different situations call for different approaches. For a quick analysis, the qualitative approach may be preferred; in other situations a mixture of qualitative and quantitative approaches may be used. In the above analysis we looked at a qualitative score method and two quantitative methods; subjective probabilities and imprecise probabilities.

The qualitative approach is crude, based on more or less precise definitions of scores. Nonetheless, it can be a useful tool for quickly pointing at the most important sources of the model uncertainties and then in its turn, it gives a basis for decision on whether the model should be accredited or not.

Both subjective probabilities and imprecision intervals were used in the quantitative analysis. The information at hand seemed to be best reflected in a single number - assigning intervals became difficult to justify. The analysis group had problems in explaining their meaning and this affected the specification process. For the practical decision making, in this case it seemed feasible to use subjective probabilities, and not much was gained by adding an additional dimension through intervals. Rather the intervals introduced some additional analysis time and led to interpretational confusion. However, in other cases we cannot exclude that imprecise probabilities could be preferred over the more common subjective probabilities due to reluctance by the experts to assign exact probabilities and the knowledge they possess may corresponded better with intervals.

The assessments are performed in step 3 and include eliciting all sources of uncertainties and assigning/judging their uncertainty magnitude with the selected approaches in step 2. Uncertainties are often related to assumptions and it is of high importance to identify them all, especially the ones that are at first hidden. Any assumption not identified is not analyzed and can in the last instance lead to undesired surprises in the future.

In step 4, decisions are made on accreditation or remodeling by judging the results of the model uncertainty assessment in light of the purpose of the model. In our case, the scope of the analysis was related to environmental concerns of a relatively small scale and more uncertainty could in such cases be allowed than if human lives were on stake. If remodeling is found necessary, one loop back to step 1 and repeats the steps until the model is judged eligible for accreditation.

As for step 5, remodeling was necessary in the case studied. Remodeling benefits from the knowledge obtained through the analysis steps 1 through 3. The structured analysis gives clear indications on where focus is to be set when proceeding, i.e. where the uncertainties are too large. In turn, this is where we must expand our knowledge to the necessary detail, resulting in a new or modified model. In last instance, the enhanced knowledge level should also be reflected in the uncertainty descriptions. In our case, the focus was set on assumption c), initiating a quest for price statistics, resulting in a new model that embeds the information found, on a sufficient detail level (NHPP with $\lambda(t)$), and uncertainty descriptions displaying smaller uncertainty.

From the case study analysis it is clear that the uncertainty relates to the condition on stationarity and the constant parameter λ , but the 'Poisson property' remains. If we were to have big uncertainties of assumption b) or d) kind, then the independence condition would have been violated and we would need to resort to other types of counting processes (cf. Aven and Jensen 1999).

6. CONCLUSIONS

This paper has demonstrated and discussed key features of a conceptual framework for analyzing model uncertainty in models used in risk assessments when no experimental data is available. Through a probabilistic model, we have shown an implementation where concepts are identified and used in structured uncertainty analysis, which in turn gives valuable input to decisions on accreditation/remodeling.

Several approaches for describing the model uncertainties are applicable, and in practice it could be most informative to use combinations of them, i.e. both qualitative and quantitative approaches.

In case of a remodeling decision, the case analyzed shows that we benefit from the antecedent analysis in recurring steps, providing focus areas for further knowledge expansion attempting to mitigate the epistemic uncertainties. The gained knowledge is embedded into a modified model with less uncertainty.

Acknowledgments

The authors are grateful to three anonymous reviewers for their useful comments and suggestions to earlier versions of this paper that have helped to improve it significantly.

Appendix A. Guidelines for providing scores in the qualitative approach for assessing model uncertainties

Based on Flage and Aven (2009), see also Selvik and Aven (2011).

Aspect	Score	Interpretation
Uncertainty	High	One or more of the following conditions are met: <ul style="list-style-type: none">- The assumptions made represent strong simplifications.- Data are not available, or are unreliable.- There is lack of agreement/consensus among experts.- The phenomena involved are not well understood; models are non-existent or known/believed to give poor predictions
	Medium	Conditions between those characterizing low and high uncertainty.
	Low	One or more of the following conditions are met: <ul style="list-style-type: none">- The assumptions made are seen as very reasonable.- Much reliable data are available.- There is broad agreement/consensus among experts.- The phenomena involved are well understood; the models used are known to give predictions with the required accuracy.
Sensitivity	High	Relatively small changes in base case values needed to bring about altered conclusions.
	Medium	Relatively large changes in base case values needed to bring about altered conclusions.
	Low	Unrealistically large changes in base case values needed to bring about altered conclusions.
Importance	High/Medium/ Low	Average of the other two aspect scores.

References

- Apostolakis, G. (1990) The concept of probability in safety assessments of technological systems. *Science*, 250, 1359-1364.
- Aven, T and Jensen, U. (1999) *Stochastic Model in Reliability*. Springer: New York.
- Aven, T. and Zio, E. (2013) Model output uncertainty in risk assessment. *International Journal of Performability Engineering*, 9(5), 475-486.
- Baraldi, P. and Zio, E. (2010) A comparison between probabilistic and Dempster-Shafer theory approaches to

- model uncertainty analysis in the performance assessment of radioactive waste repositories. *Risk Analysis*, 30(7), 1139-1156.
- Bjerga, T., Aven T. and Zio, E. (2012) An application of a new framework for model (output) uncertainty analysis in risk assessment. *Proceedings of PSAM'11/ESREL 2012. 25-29 Juni 2012, Helsinki*.
- Bayarri M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.H., and Tu, J. (2007) A framework for Validation of Computer Models. *Technometrics*. 49(2), 138-154.
- de Rocquigny, E., Devictor, N. and Tarantola, S. (2008) *Uncertainty in Industrial Practice. A guide to quantitative uncertainty management*. Wiley: NJ.
- Devooght, J. (1998) Model uncertainty and model inaccuracy. *Reliability Engineering and System Safety*, 59, 171-185.
- Droguett, E.L. and Mosleh, A. (2008) Bayesian Methodology for Model Uncertainty Using Model Performance Data. *Risk Analysis*, 28(5), 1457-1476.
- Droguett, E.L. and Mosleh, A. (2013) Bayesian Treatment of Model Uncertainty for Partially Applicable Models. *Risk Analysis*. Accepted for publication.
- Flage, R. and Aven, T. (2009) Expressing and communicating uncertainty in relation to quantitative risk analysis (QRA) *Reliability & Risk Analysis: Theory & Applications*. 2(13), 9-18.
- Helton, J.C. (1994) Treatment of uncertainty in performance assessments for complex systems. *Risk Analysis*, 14, 483-511.
- Helton, J.C., Johnson, J.D. and Oberkampf, W.L. (2004) An exploration of alternative approaches to the representation of uncertainty in model predictions. *Reliability Engineering and System Safety*, 59, 171-185.
- Jiang X., Yang, R. J., Barbat, S. And Weerappuli, P. (2009) Bayesian Probabilistic PCA Approach for Model Validation of Dynamic Systems. *Proceedings of SAE World Congress and Exhibition*, Detroit, MI, April 2009.
- Kaminski Jr., Riera, J.D., de Menezes, R.C.R., Miguel, L.F.F. (2008) Model uncertainty in the assessment of transmission line towers subjected to cable rupture. *Engineering Structures*, 30, 2935–2944.
- Kennedy, M. C. and O'Hagan, A. (2001) Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 63(3), 425-464.
- Knupp, P. (2002) *Verification of Computer Codes in Computational Science and Engineering*. Chapman & Hall/CRC: Boca Raton, FL.
- Lindley, D. V. (2000) The philosophy of statistics. *The Statistician*, 49, 293-337.
- Lindley, D.V. (2006) *Understanding Uncertainty*. Wiley: Hoboken, NJ.
- McFarland, J. M. (2008) *Uncertainty Analysis for Computer Simulations through Validation and Calibration*, PhD Thesis, Vanderbilt University.
- Meeker, W.O. and Escobar, L.A. (1998) *Statistical Methods for Reliability Data*. Wiley:N.Y.
- Nilsen, T. and Aven, T. (2003) Models and model uncertainty in the context of risk analysis. *Reliability Engineering and System Safety*, 79, 309-317.

- Oberkampf, W. and Trucano, T. (2002) *Verification and Validation in Computational Fluid Dynamics*. Technical Report SAND2002-0529, Sandia National Laboratories, Albuquerque, New Mexico.
- Östergaard, C., Dogliani, M., Guedes Soares, C., Parmentier, G. and Pedersen, P.T. (1996) Measures of model uncertainty in the assessment of primary stresses in ship structures. *Marine Structures*, 9, 427-447.
- Parry, G. and Winter, P.W. (1981) Characterization and evaluation of uncertainty in probabilistic risk analysis. *Nuclear Safety*, 22(1),28-42.
- Parry, G. and Drouin, M.T. (2009) Risk-Informed Regulatory Decision-Making at the U.S. NRC: Dealing with model uncertainty. Nuclear Regulatory Commission, 2009.
- Rebba, R., Mahadevan, S., and Huang, S. (2006) Validation and Error Estimation of Computational Models. *Reliability Engineering and System Safety*, 91,1390–1397.
- Reinert, J.M and Apostolakis, G.E. (2006) Including model uncertainty in risk-informed decision making, *Annals of Nuclear Energy*, 33(4) 354-369.
- Roache, P. (1998) *Verification and Validation in Computational Science and Engineering*. Hermosa Publishers: Albuquerque, NM.
- Rosqvist, T. and Tuominen, R. (2004a) Precautionary risk decision-making. PSAM7 & ESREL'04 International Conference on Probabilistic Safety Assessment and Management, Berlin, 14-18 June 2004. Springer-Verlag: London, pp. 567 –572.
- Rosqvist, T. and Tuominen, R. (2004b) Qualification of Formal Safety Assessment – An Exploratory Study. *Safety Science*, 42(2), 99–120.
- Selvik, J.T. and Aven, T. (2011) A framework for Reliability and Risk Centered Maintenance. *Reliability Engineering and System Safety*. 96, 324-331.
- US Nuclear Regulatory Commission. *Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making* (NUREG-1855). Washington DC, 2009.
- US Nuclear Regulatory Commission. *Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making Draft Report for Comments* (NUREG-1855, rev. 1). Washington DC, 2013.
- Vasseur, D., Le Duy, T.D., Dutfoy, A., Dieulle, L. and B érenguer, C. (2012) Uncertainty analysis in probabilistic risk assessment: Comparison of probabilistic and non probabilistic approaches. In: *Advances in Safety, Reliability and Risk Management: ESREL 2011*, Troyes, (eds. B érenguer, C., Grall, A., Soares, C. G.), Taylor & Francis Group.
- Xiong, Y., Chen, W., Tsui, K.L., and Apley, D. W. (2009) A Better Understanding of Model Updating Strategies in Validating Engineering Models. *Journal of Computer Methods in Applied Mechanics and Engineering*, vol. 198, no. 15-16, pp. 1327-1337.
- Zio, E. (2006) A Study of the Bootstrap Method for Estimating the Accuracy of Artificial Neural Networks in Predicting Nuclear Transient Processes, *IEEE Trans. On Nucl. Sci.*, Vol. 53, no. 3, pp. 1460-1478.
- Zio, E. and Apostolakis G.E. (1996) Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories. *Reliability Engineering and System Safety*, 54, 225-241.